

2. félévi beszámoló

Bircz András Mátyás (abircz@caesar.elte.hu)

Statisztikus Fizika, Biológiai Fizika és Kvantumrendszerek Fizikája PhD program

Témavezető: Csabai István

A dolgozat címe: Gépi tanulás a tudományokban

2020.06.18.

Bevezetés:

Az előző féléves doktori beszámolómban leírt bevezető jelenleg is aktuális, a kutatási témám nem változott. A mesterséges intelligencia egyre gyorsabb, egyre szélesebb körben való elterjedését segíti, hogy lehetőséget nyújt komplex, adatintenzív feladatok, problémák megoldására. Nem feltétlenül cél az adott problémakör teljes megértése, leginkább az számít, hogy az adott módszer minél pontosabban működjön. Az esetek többségében nincs is lehetőség a modellek részletes megértésére, mivel ezek általában milliós (akár tíz- vagy százmilliós) nagyságrendű paraméterrel rendelkeznek.

Az aktuális félévben elvégzett kutatások ismertetése:

A félév során folytattam az előzőben elkezdett projektet, mely az El-Nino előrejelzésének fejlesztéséről szól gépi tanulás és mélytanulási algoritmusok segítségével. A Nature-ben megjelent cikk [1] alapján haladva saját feldolgozó kódokkal megfelelően előkészítettem a bemeneti adatokat a tanításhoz. Ezalatt többször diszkutáltam a témavezetőmmel ennek a folyamatnak a mikéntjéről, mivel a kiindulásnak tekintett cikkben nem tértek ki erre. A saját kódok írásával párhuzamosan többször kísérletet tettem a cikk szerzőivel való kommunikációra, mely sajnos végül nem vezetett a kellő eredményre, nem sikerült megtudni elég információt, alig kaptam választ a feltett kérdéseimre. Ennek ellenére sikerült elvégezni az adatok feldolgozását, további adathalmazok felkutatását, a tanítóhalmaz javítását. Többször ellenőriztem a kódjaimat, irodalomkutatást is végeztem, amely során találtam néhány cikket, melyek különösen is a segítségemre voltak [2][3]. Végül sajnos nem sikerült reprodukálnom a kiindulásnak vett cikket [1], de sikerült egy működőképes programcsomagot elkészítenem, mely github-on megtalálható¹. Ezalatt sokat tanultam, leginkább a klimatológiai adatok feldolgozásáról (pl. netCDF formátum), de a neurális háló könyvtárak használatáról is (pl. Tensorflow, Keras).

A vizsgálataim és a szakirodalom alapján a témavezetőmmel arra a következtetésre jutottunk, hogy a cikk szerzői túlságosan optimisták voltak a saját eredményeik megítélésében. Valószínűleg, valamilyen nem kellően dokumentált hiba miatt átszivároghatott információ a tesztelésbe, ezért jöttek ki meglepően jó eredmények, ezért nem sikerült rekonstruálni a szerzőkkel való többszörös konzultáció után se. Úgy döntöttünk, hogy túl sok munka lenne alaposan felderíteni és alátámasztani a hibákat, úgyhogy egyelőre ennek a témának a további vizsgálatát félretesszük.

A félév vége felé két új témakörben is elindultak kutatások, melyekben részt fogok venni: egyrészt Oroszlány Lászlóval és Barcza Gergellyel közösen arra teszünk kísérletet, hogy gépi tanulási módszerekkel segítsük a kvantummechanikai soktest problémák számolását, másrészt a Népegészségügyi Központtal kezdtünk meg kutatási egyeztetéseket, ahol pollenek gépi tanulással segített felismerése lenne a cél.

1 https://github.com/abircz/ENSO_DL

Publikációk:

Ebben a félévben közel kerültem egy, az El-Nino előrejezés fejlesztéséről szóló cikk írásához, de mint ahogy már említettem a fenti okok miatt ez nem jött össze.

Tanulmányi tevékenység az aktuális félévben:

Összesen 2 kurzust vettem fel és végzek el, melyek az alábbiak:

-Elméleti evolúcióbiológia EA (FIZ/3/005E)

-Sejtszignalizációs hálózatok kvantitatív analízise (FIZ/3/055E)

Konferenciák, workshopok az aktuális félévben:

Rendszeresen részt vettem a tanszéken csütörtökönként tartott olvasószemináriumon, melyen a mesterséges intelligenciával kapcsolatos érdekességekkel, újdonságokkal foglalkoztunk. A félév során kialakult vírushelyzet miatt nem tudtam részt venni személyesen konferenciákon, workshopokon, remélhetőleg a következő félévtől újból lesz erre lehetőség.

Oktatási tevékenység az aktuális félévben:

A félév során az alábbi két tárgy oktatásában vállaltam szerepet, melyeket gond nélkül sikerült távoktatásban folytatni:

-A fizika numerikus módszerei (fiznum1f19la, labor, 3 kredit) gyakorlat egyik vezetője voltam.

-Tudományos modellezés számítógépes laboratórium (dsscimodf17lm, labor, 5 kredit) egyik témavezetője voltam, egy hallgató projektmunkáját kísértem végig a félév során.

Ezeken kívül még egy fizika alapszakos hallgató témavezetésében segédkeztem a félév elején.

Hivatkozások:

[1] Ham, Y., Kim, J. & Luo, J. Deep learning for multi-year ENSO forecasts. Nature 573, 568–572 (2019). <https://doi.org/10.1038/s41586-019-1559-7>.

[2] Saha, M, Nanjundiah, RS. Prediction of the ENSO and EQUINOO indices during June–September using a deep learning method. Meteorol Appl. 2020; 27:e1826. <https://doi.org/10.1002/met.1826>

[3] He, Dandan & Lin, Pengfei & Liu, Hailong & Ding, Lei & Jiang, Jinrong. (2019). DLENZO: A Deep Learning ENSO Forecasting Model. 10.1007/978-3-030-29911-8_2.